DOCUMENT RESUME

ED 192 278 CS 005 652

AUTHOR Canfield, Robert
TITLE A Study of the Reliability and Selected Aspects of Validity of a Single Time-Line Format for Coding Verbal Interaction During Reading Instruction.
INSTITUTION State Univ. of New York, Oswego. Coll. at Oswego.
SPONS AGENCY State Univ. of New York Research Foundation, Oswego.
PUB DATE Dec 79
NOTE 17p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Research: Elementary Education: *Evaluation Methods: *Interaction: *Interaction Process Analysis: *Reading Instruction: Reading Research: Reliability: *Student Teacher Relationship: Teacher Effectiveness: Validity

ABSTRACT

Prompted by the need to develop improved tools for describing and assessing teachers' skills in teaching reading, a study was undertaken to examine the effectiveness of a system for classifying and tallying teacher-pupil verbal interaction during classroom instruction. Specifically, the study investigated the extent of agreement among college students using a time-line format when ranking categories of verbal interaction in three criterion reading lessons, and it examined the use of selected ratios of verbal interaction as a means of discriminating among criterion reading lessons as compared to the judgments of a jury of reading teachers who ranked the quality of the lessons without reference to the verbal interaction factors. Twenty-five graduate students were trained in the Flanders Interaction Analysis technique using the time-line format. The students then coded the verbal interaction of three videotaped reading lessons. These lessons were also viewed by seven reading teachers, who ranked them according to overall effectiveness. The results indicated that the students reached high agreement in tallying the verbal interaction. In addition, the use of the ratio of teacher questions/pupil responses to teacher lecture/pupil initiated comments was found to rank the lessons in the same quality order as did the jury of reading teachers. (Author/FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED1922278

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A Study of the Reliability and Selected Aspects of Validity of a Single Time-Line Format for Coding Verbal Interaction During Reading Instruction

Robert Canfield
Investigator

State University College at Oswego

December 1979

Sponsor

University Awards Committee
Research Foundation
State University of New York
Account #228-7150A

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Robert Canfield

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS005652

Purpose

The purpose of this study was to examine the effectiveness of a system for classifying and tallying the teacher-pupil verbal interaction during classroom instruction in reading. Growing out of a need to develop improved tools for describing and assessing a teacher's professional skills in teaching reading, the investigator undertook the task of examining the reliability of college students' use of a single time-line format for coding classroom verbal interaction using Flander's ten categories of verbal interaction. (1970)

While the main purpose of this study was to examine college students' degree of agreement using the time-line format, the design of the study allowed for the collection of certain descriptive information that also reflected on the validity of the technique as a measure of the overall quality of reading instruction. A secondary purpose, therefore, was to examine the concurrent validity of the techniques as established by the degree of agreement with a jury of reading specialists' evaluations of the overall quality of selected criterion reading lessons.

Background

From July 1974 until June 1976 the State University College at Oswego was engaged in the development of a program of pre-service education of teachers to teach reading through a grant which was funded by the United States Office of Education. (Grant #OEG-0-74-8988) An important aspect of the program related to the assessment of teaching competencies associated with teaching reading skills. Since one of the observable features of skill instruction is the verbal exchanges that occur between the teacher and pupils, an attempt was made in the project to develop techniques for measuring verbal interaction during reading lessons. Recognizing that measures of verbal interaction would be descriptors of events, rather than assessments of quality of the events, the

investigators made no claim that measures of verbal interaction, used singly, would be useful in yielding scores that would meet necessary criteria in judging reading instruction. However, there was the possibility that certain patterns of verbal interaction, combined with descriptors of other classroom variables, would be useful in the ultimate assessment of teaching competency. This possibility prompted the investigators to work with the verbal interaction technique. As a result, a system for tallying verbal interaction in the context of reading lessons was designed using Flander's ten categories of verbal interaction which resulted in identification of certain verbal interaction ratios that appeared to be sensitive to the quality of reading lessons. For example, certain ratios from the adapted version of the Flander's categories when applied to the college students' verbal interaction tallies ranked the five criterion reading lessons in the same ascending quality order as previously validated observation checklists.

There was evidence, however, that college students needed more training in coding verbal interaction to provide reliable measures of the verbal exchanges in reading skills lessons. It was clear that further changes in the format used to tally verbal interaction were needed to allow the format to be used with the variety of reading skill lessons currently employed in the schools. The current study, then, was planned to redesign the format for coding verbal interaction and investigate the validity and reliability of the new instrument.

The Problem

Specifically, the purpose of this investigation was to determine the degree of agreement among college students using the time-line format when ranking categories of verbal interaction in three criterion reading lessons. Information was also sought regarding the use of selected ratios of verbal interaction

in discriminating among criterion reading lessons when compared to the judgments of a jury of reading experts who ranked the quality of the lessons globally, i.e., without direct reference to verbal interaction factors per se.

The following questions guided the study of the problem:

1. Will first year graduate students, after completing a six-hour training module with the time-line version of the Flander's Interaction Analysis technique, reach significant levels of agreement in ranking categories of verbal interaction when analyzing three video tapes of reading skills lessons?
2. Will ratios be discernable among the tallies in the verbal interaction categories that will sequence the three reading skills lessons in the same order of quality as that indicated by the jury of reading experts?

Procedure

1. A horizontal time-line format was prepared for use in reading tallies during skill instruction in reading. The time-line developed is similar to Urback's "interaction sequence graph" as reported in Flander's Analyzing Teaching Behavior. (1970) The time-line format was trial tested as advanced graduate students taught skills lessons as part of their reading practicum course.

2. A six-hour module was prepared to train first year graduate students in the techniques of tallying and classifying verbal interaction. The module makes heavy use of Kryspin and Feldhusen's volume, Analyzing Verbal Classroom Interaction. (1974) Audio and video tapes of reading lessons were prepared to be included in the activities of the module.

3. Three versions of a lesson designed to teach intermediate grade pupils

how to select the correct meaning of a word from a dictionary entry were prepared. Lesson plans for each lesson were prepared with approaches to the skill altered somewhat in each case -- but with the objectives for the pupils remaining the same. The lessons were taught by the same certified teacher -- each lesson to a different group of approximately eight pupils. No direct attempt was made to differentiate verbal interaction patterns in the lessons other than those differences that naturally occurred in carrying out the particular skill approach being used. The lessons were video tape recorded and later ranked by the teacher and the investigator in terms of their overall effectiveness in meeting the pupils' objectives.

4. A jury of seven certified reading teachers currently employed in local public schools viewed the video tape recordings of the three criterion lessons. This jury was directed to simply rank the lessons in terms of each lesson's overall effectiveness in developing the skill objective. After ranking the lessons, jurists were asked to write comments on why they ranked the lessons as they did. Jurists were not informed that verbal interaction was the main interest of the investigators. The jury had complete agreement of their ranking of the lessons and concurred with the rankings of the investigator.

The following comments are samples of the reasons given by the jurists for their ordering of the tapes:

Tape A	Teacher "dominated"	Did not draw meanings (of words) from pupils
	"No interaction"	Little enthusiasm
	"Teacher redundant"	"Little student involvement"
	"Did not ask pupils"	"Not enough language output from children in their "own" words"
	"Did not ask if all agree"	
	"Teacher told what was learned -- did not ask"	

Tape B

Some positive reinforcement (good)
Pupil mistakes not looked down upon
Not as much rephrasing of the pupils' responses
"Teacher answered her own questions"
"Complimented kids -- maybe too much"
T-asked for definitions where they were not important
"Group appeared not to be zeroed in on the task"
"Children who could shout the loudest got to answer"
"Teacher verbalized with boys to keep them under control"
"Always more beneficial to ask students rather than tell all terms (definitions). Then they have to think it out."
"There was good children involvement although a few children dominated the lesson"
"Children's response were livelier (than A) and sometimes inappropriate."
"Lacked organization"
"T-didn't establish correctness of responses"
"Good to get summary from kids"
"Boys were hyper and monopolized the lesson"

Tape C

"More verbal (2) positive reinforcement some smiling"
"More feeling shown for the child" (1)
"Explicit and positive" (2)
"Highly accepting" (1)
"Asked 'how' questions"
"Asked if pupils agree before going on to next item"
"Very good summary of lesson and teacher could evaluate if the children understood the skills being taught"
"Child centered with good guidance by teacher"
"Has class participation"
"Good pupil involvement"
"Teacher sat down with students -- closer contact"
"Built upon student responses" (3)
"Used discovery method" (3)
"Asked pupils to state rule generalization"
"Questions zero in on helping children understand the process they're using"
"Asked them (pupils) to tell why they chose a particular response"

5. Twenty-five first year graduate students, enrolled in R. Ed. 501, Foundations of Reading, in the Summer of 1978, were trained in the Flander's technique for recording verbal interaction, using the previously discussed module. After their training, these students used the time-line format to record the verbal interaction that occurred during the three criterion lessons that had been previously video taped. A section of 19 graduate students, enrolled in the same course of the fall semester of 1978, followed

the same procedures. A total of 44 graduate students, therefore, used the time-line format in recording verbal interaction of the criterion tapes.

6. Data from the graduate students' tallying of the verbal interaction of the three criterion lessons allowed use of the Kendall Coefficient of Concordance to ascertain the degree of agreement among the students in ranking categories of verbal interaction. Kendall's W is a technique used to express degree of agreement among a series of judges (in this case graduate students) ranking a series of events (in this case categories of verbal interaction).

$$W = \frac{S}{\frac{1}{12} K^2 \quad (N^3 - N)}$$

Use of Kendall's Concordance yields values that are similar in appearance and interpretation to a correlation coefficient and indicate degrees of consistency of the completed rankings among raters.

7. Empirical inspection of the relationships among categories of verbal interaction was necessary to identify ratios, or combination of ratios, that would possibly provide a single index that would be sensitive to the differences in the verbal interaction in the three criterion lessons. If use of this index placed the three criterion tapes in the same order as the jury of experts, then there would be support for establishing this index as a measure of the quality of skills lessons in reading. Further, if the identification and weighting of factors used in computing the index tended to agree with the reasons that the jury of experts used ranking the lessons, further support of validity would be in evidence.

Limitations

1. Reading instruction can include many types of lessons depending on the particular aspect of the reading curriculum to be emphasized. A skill lesson is just one of these types, and, depending on its objectives, can be

organized in different ways. The type of skill lesson employed in this study for all three criterion lessons was adapted from the Houghton-Mifflin Company's Reading for Meaning Series - McKee (et.al.), 1970. Components of this type of skill lesson include: (1) teacher explanation; (2) teacher-led group practice; (3) pupil individual practice; (4) evaluation and summary. The results of this study are limited by the degree that this format is truly representative of skill lessons, generally. It is also limited by the extent to which the particular skill instructed in the criterion tapes is truly representative of skill topics, generally.

2. Three criterion skill lessons were prepared and used in this study. This number of skill lessons clearly restricted the options available to the jury in identifying lessons of contrasting quality. Results of the study, therefore, are limited to the extent that the three versions of skill lessons are reflective of "quality" differences that exist in reading skill lessons generally.

3. Critical to the design of the study are the ratings of the quality indicated by a jury of "reading specialists". The jury in this study was composed of seven reading specialists with an average of ten years of classroom experience and an additional five years of experience as reading specialists in public schools. All possess graduate credits in reading with six holding Master's degrees. All are certified teachers in New York State. The study is limited, however, to the extent that these jurists truly reflect accurate perceptions of quality of reading instruction.

Results

1. Information was sought regarding the level of agreement among graduate students in ranking Flander's ten categories of verbal interaction that occurred during the three criterion skills lessons. Kendall's Coefficient Concordance

was used for this purpose.

With 19 as the minimum number of judges (graduate students) involved in the ranking of the ten categories in any one of the six presentations, a value for concordance exceeding .16 would be considered significant at the .001 level. Table I indicates the concordance values for three taped lessons when tallied by graduate students enrolled in both the Summer and Fall semesters.

TABLE 1

Values of Concordance for Three Tapes at Two Collection Points

Tape	Summer 1978 Judges	Summer 1978 W*	Fall 1978 Judges	Fall 1978 W*
A	22	.857	19	.907
B	25	.770	19	.849
C	25	.865	19	.865

*Corrected for Ties

Inasmuch as the smallest value of W achieved was over .75, it is clear that ranks formed on the basis of observations of categorized events are in high agreement across the three tapes and across the two data collections.

The use of concordance and the resulting values indicates that judges trained in this type of observation technique can and do consistently make a series of observations which when ranked in order of occurrence are highly consistent with others in the rating group. Moreover, significant agreement is found across a range of three skill lessons identified by a jury of reading experts as representing three different quality levels of instruction.

2. While significant agreement was found among the college students' rankings of ten categories of verbal interaction, use of these rankings for each lesson failed to produce a pattern that would sequence the lessons in the

order indicated by the jury of reading specialists. Apparently, sensitivity to the quality differences in the three criterion lessons became diffused when the tallies in each category were converted to single rank values.

3. Empirical inspection of the relationships among number of tallies in each category (rather than the rank order of the category) did reveal certain patterns of verbal interaction that appear to discriminate among the three criterion tapes. Table II presents the mean percentage of tallies in each category for each group of coders.

TABLE II

Mean Percentage of Tallies for the Ten Categories of Verbal Interaction in the Criterion Lessons as Coded by Two Groups of College Students

Category	Lesson A		Lesson B		Lesson C	
	Group I(S.S.)	Group II(Fall)	Group I	Group II	Group I	Group II
1	0	0	0	0	0	0
2	2	3	3	4	5	6
9	1	0	25	30	6	7
3	4	2	5	4	4	4
5	62	63	10	11	20	19
4	15	16	19	18	26	26
8	8	8	10	11	18	17
0	5	6	20	20	15	16
6	5	4	7	5	6	4
7	0	0	0	0	0	0

Inspection of Table II reveals that there is relatively high agreement among the coders that lesson A is characterized by much teacher lecture, while Lesson B

is unique in its high incidence of pupil initiated comment. Lesson C tends to reflect a middle ground in regard to teacher lecture and pupil initiated comments, but shows somewhat higher pattern of teacher questioning and pupil response. Also, while the number of tallies in these categories is quite limited, Lesson C reflects higher percentage of teacher praise (category 2) than either Lesson A or B.

In order to identify any existing pattern of verbal interaction that would rank the tapes in the same order as the jury of reading teachers, the investigator explored the ratios created by various combinations of the mean percentages of the high frequency categories. Based on this review, it was found that the sum of the mean percentage of tallies in category four and eight (teacher questions and pupil responses) divided by the sum of the mean percentage of tallies in five and nine (teacher lecture and pupil initiated comments) yielded ratio values that ranked the criterion lessons in the same order as the jury of reading specialists. With this information as a guide, each graduate students' $\frac{4+8}{5+9}$ ratio was computed for the three criterion lessons to determine the percentage of coders who ranked the lessons in the "correct" order. Table III presents the means and standard deviations for each group of college students using this particular ratio.

TABLE III

Means and Standard Deviations for the $\frac{4+8}{5+9}$ Ratio
for Two Groups of College Student Coders

	Lesson A		Lesson B		Lesson C	
	M.	S.D.	M.	S.D.	M.	S.D.
Group I n=19	.37	.09	.76	.61	1.87	.83
Group II n=21	.37	.12	.98	.71	1.95	.82

An examination of the percent of coders whose tallies would rank the lessons in the "correct" order when using the $\frac{4+8}{5+9}$ ratio was computed. To establish the values of $\frac{4+8}{5+9}$ that would be the critical points on a scale between the A, B, and C lessons, an assignment of values was made based on maximizing the number of "hits" (coders whose $\frac{4+8}{5+9}$ ratios ranked each lesson in its correct position in regard to the jury's quality order) for the group II coders. Values assigned on this basis were .455 between Lesson A and B and 1.245 between Lesson B and C.

Eighty percent of the group II coders, using this ratio, ranked the lessons in the correct order. Using these same critical scale points, 79 percent of the group I coders also ranked the lessons in the "correct" order. It is apparent, then, that this ratio of verbal interaction categories will tend to rank lessons (approximately four out of five times) in the same order as a jury of reading specialists who use overall lesson quality as their major criterion. Further, the closeness of the percentage of "hits" in each independent group of coders tends to further support the relatively high degree of agreement among the coders in tallying verbal interaction. Table IV presents the percentage of "hits" for each criterion lesson for each coder group.

TABLE IV

Percentage of Coders Placing Lessons in "Correct" Quality Order Using $\frac{4+8}{5+9}$ Ratio of Flanders' Categories

	Group I	Group II	Combined Groups
Lesson A	76	89	82.5
Lesson B	86	63	75
Lesson C	76	89	82.5
Total	79	80	79.5

Inspection of the reasons given by the jury in rating Lesson "A" as "poorest" and Lesson "C" as "best" also points to the critical nature of the ratio of teacher questions and pupils answers to the amount of teacher lecture and pupil initiated comments. Jurists' comments related to tape A included: "Teacher did not ask questions,", "Teacher told what was learned -- did not ask.", and "Teacher did not draw meanings of words from pupils". Where as in regard to tape C, the comments included: "The teacher asked "how" questions,", "The teacher asked if pupils agree before going on to the next item.", and "Asked pupils to state rule generalization."

Conclusions

1. College students use of a single time-line format in tallying verbal interaction is apparently a reliable measure of the verbal exchanges during skill instruction in reading. The amount of agreement among the college student raters, 80%, indicates a level of success in the use of the technique that suggests acceptable levels of agreement among student raters.

2. When based on the interpretations of quality (as reflected in a jury's perception of quality in the lessons used in this study) the use of at least one selected ratio of verbal interaction categories $\left(\frac{4+8}{5+9}\right)$ will rank the lessons in quality order. The incidence of teacher questions and pupil answers, expressed in units of teacher lecture combined with pupil initiated comments, results in the desired quality ranking. There is an indication, therefore, that the number of questions and answers in comparison to the amount of teacher lecture and other pupil comments is a valid measure of quality among lessons similar to the criterion lessons used in this study. Further study of this possible relationship is needed, especially as it relates to the upper limits of the values yielded by this ratio. It is entirely possible, for example, that at some point the proportion of questions and answers in a reading skill

lesson is no longer a useful index to the quality of the lesson. This study did not establish this upper limit. However, assuming that most reading skill lessons possess proportions of questions and answers that are comparable to the proportions found in this study, the index appears useful. Further study is also needed to ascertain if lessons aimed at teaching contrasting types of skills require different proportions of questions and answers. The ease and reliability with which raters can consistently tally these particular categories of verbal interaction recommends the ratio for use in teacher education settings. Both the verbal interaction patterns and the written comments of jurists point to the importance of this particular ratio.

3. Inspection of various categories of verbal interaction in the three criterion lessons also illustrates that large percentages of teacher talk, as found in Lesson A, or large amounts of pupil initiated comments, as found in Lesson B, are not perceived by a jury as indicators of high quality. Rather, patterns that reflect modest amounts of these factors are seen to be associated with the lesson that rates the highest in quality (Lesson C). This suggests skill instruction in reading needs to achieve a middle ground between lessons that contain large amounts of teacher talk and lessons that have large amounts of pupil talk.

Additional Comments

In addition to the previously stated results, analysis of the data from this investigation suggested certain relationships between skill instruction in reading and verbal interaction that are worthy of note. The following statements, while not conclusively established by the analysis of the data, are listed here for purposes of future study:

1. When comparing the verbal interaction of the (jury - perceived) poorest lesson to the best lesson, larger amounts of teacher acceptance,

teacher praise and teacher building upon pupil responses were found in the best lesson. However, because of the low frequency of tallies in these categories in all three of the lessons, these factors did not discriminate statistically among the three criterion lessons. It is recommended that Flander's Teacher Response Ratio, Teacher Question Ratio and Pupil Initiation Ratio be used to help overcome this low frequency problem. (Flanders, 1970, pg. 102)

2. The jury tended to recognize pupil initiated comments as a desireable quality of the skill lessons, but reacted negatively when they felt these comments were "off task" or in excess. This was observable in the jury's ranking of the lessons and in their written comments about the criterion tapes. This suggests that there is an optimum level or range of pupil initiated comments that needs to be sought, but not surpassed, in reading skill lessons of the type used in this study.

3. Study of verbal interaction in the various functional parts of a skill lesson (explanation, group practice, independent practice, etc.) did not reveal greatly different patterns from those of the total lesson. Study of verbal interaction patterns of whole lesson rather than parts of lessons is suggested. Analysis of verbal interaction patterns of skill lessons compared to directed reading activities as well as language experience lessons would indicate if verbal interaction patterns were a function of the type of reading lesson employed.

REFERENCES

Canfield, Robert (Project Director), Right to Read Grant #OEG-0-74-8988 A Program of Pre-Service Education of Teachers to Teach Reading. Final Report, 1976.

Flanders, Ned A. Analyzing Teaching Behavior. Reading, Pennsylvania: Addison-Wesley Publishing Company, 1970.

Kryspin, William J. and Feldhusen, John F. Analyzing Verbal Classroom Interaction. Minneapolis, Minnesota: Burgess Publishing Company, 1974.